**CMPEN 454**

**Project 1**

**Fall 2016**

Created by: Andrew Lee

Nathan Scribano

Eric Traister

1. Summary

The convolutional neural net (CNN) project was a first step into understanding fundamentals of modern computer image processing design. Implementing a forward pass CNN was a deep learning experience that stretched our understanding of object recognition from an algorithmic standpoint. The project as a whole was split into several stages of development from structural conception to layer by layer development and finally merge and design verification.

With the assistance of a debugging interface, development was easily verified in each layer of design allowing us insight into expected results. Skills attributed to and learned from in the CNN project design included many Matlab programming techniques and algorithms. Most of our team had not extensively used Matlab in the past, and as a result, we gained great experience working with the powerful language in extensive use of general multi-dimensional arrays and specific image processing and filtering techniques.

The quantitative performance evaluation gave us insight into validating and verifying the functionality of our simple CNN. We did not have any initial expectations of the neural net’s performance so the results obtained were discussed as plausible given the sophistication and depth of the CNN.

1. Outline – procedural approach, flowchart of structure

Initial development of the CNN was focused towards creation of a working CNN from the image set extracted from the ‘cifar10testdata’ data set. The procedural approach started with a function that loaded the test data, neural net filters and parameters, and the provided debugging test data for design verification.

We created separate demo functionality in the CNN to further verify the efficiency and performance of the neural net. Input images are read from the respective directories the demo is contained in. These images are then converted into working Matlab test data and pushed through a modified version of our CNN tweaked to evaluate each inputted image individually with intermediate results showing progress of “learned” features.

In our main neural net function, the image set extracted from the test data set is sized and parsed in a for loop for each image the set contains. After converting the current image to double, the image then forward prorogates through the neural net layer by layer. The CNN’s layers were structured accordingly:

Normalization Layer:

* Input is the current ‘image’. Output returns the image computed with the provided mathematical normalization calculation. No special design decisions required.

Convolution Layer:

* Input is the current ‘image’ and the CNN parameter’s filter banks along with the bias values. Output returns the convolved image by iterating over every value of the image’s array and convolving the image with each of the predefined linear filters and bias values.
* Design consisted of a double nested for loop that iterates over and convolves the image’s three channels into the filter bank and corresponding bias value’s ten channels.

Rectified Linear Unit:

* Input is the current ‘image’. Output returns the image computed with the provided mathematical ReLU calculation. No special design decisions required.

Maxpool Layer:

* Input is the current ‘image’. Output returns the image’s max value at each 2x2 sub-array to decrease the spatial dimensions of the image to a more manageable and smaller representation of the input image.
* Maxpool has to operate over each channel individually and does so by iterating through each channel with a for loop and chopping the image into four arrays with 2x2 sub arrays that are then compared to find top/bottom and left/right max values. These are then converged to find the overall max value of the top and bottom halves.

Fully Connected Layer:

* Input is the current ‘image’. Output returns an array similar to the convolution layer but limited to a 1x1xD array that contains only the image’s channel convolution properties.
* A quadruple nested for loop was used to iterate across each NxM pixel and each D dimension in the image along with each filter bank level that must be used for convolution. This step is the most time consuming layer as it must convolve each image with an equivalent size filter for every filter in the bank.

Softmax Layer:

* Input is the current ‘image’. Output returns a set of probabilities from 0 to 1 that resulted in the neural net’s evaluation of the input image.
* The design iterates across each dimension using a for loop and computes the output as the result given to us with the provided mathematical exponentiation equation to find the probabilities.

Result Analysis:

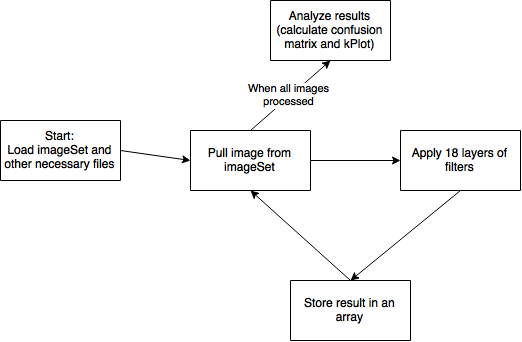
Confusion Matrix Calculation:

* + The confusion matrix is a 10x10 array, where the columns represent the correct image class and the rows represent the CNN’s guessed image class for each image in the set. The input is the results and the length of the results.
  + The function iterates across each image of the results, and adds 1 to the location of the confusion matrix where the correct class column and guessed class row intersect.
  + The output is the 10x10 array after each image’s guess is calculated.

k-Plot Calculation:

* For each 1x1x10 softmax output:
  + Create 10 index-value pairs, like a map, but stored in a cell array.
    - For example: (1, 2.6%), (2, 50.1%), … , (10, .5%)
  + Sort the pairs by value, so that the highest percentage is the first value in the cell array
  + Iterate from i = 1:10 until trueclass(i) == cellArray(i)
    - From i to end, add 1 to k-plot array (10x1)
* Divide k-plot array by number of softmax outputs and multiply by 100 to get percentages
  + 10,000 for this project.

Flowchart of Execution:



1. Experimental Observations

For the most part, the program runs as we expect it. Because many operations were vectorized, the program runs efficiently for most steps, except for fullconnect(). This is because the only way we knew to accurately implement fullconnect() was to nest a number of for loops together. As for how accurately the CNN performs, we initially thought its top guess accuracy of 43.7% seemed a little low, but after thinking about how more accurate CNNs are often composed of many more than 18 layers, we figured the accuracy isn’t too bad. When looking at the intermediate images during the process, some of the filters appear like we would expect – strongly reacting to the class’s filter – while others did not react as much as we anticipated. Intermediate images from demo1() are shown below.

Input image (32x32x3):

demo/airplane.jpg

After layer 3 (32x32x10):

demo/layer3-1.jpg demo/layer3-2.jpg demo/layer3-3.jpg demo/layer3-4.jpg demo/layer3-5.jpg demo/layer3-6.jpg demo/layer3-7.jpg demo/layer3-8.jpg demo/layer3-9.jpg demo/layer3-10.jpg

After layer 8 (16x16x10):

demo/layer8-1.jpg demo/layer8-2.jpg demo/layer8-3.jpg demo/layer8-4.jpg demo/layer8-5.jpg demo/layer8-6.jpg demo/layer8-7.jpg demo/layer8-8.jpg demo/layer8-9.jpg demo/layer8-10.jpg

After layer 13 (8x8x10):

demo/layer13-1.jpg demo/layer13-2.jpg demo/layer13-3.jpg demo/layer13-4.jpg demo/layer13-5.jpg demo/layer13-6.jpg demo/layer13-7.jpg demo/layer13-8.jpg demo/layer13-9.jpg demo/layer13-10.jpg

The output results from the softmax layer:

airplane 42.228%

automobile 1.662%

bird 0.863%

cat 0.160%

deer 0.506%

dog 0.029%

frog 0.055%

horse 0.069%

ship 53.749%

truck 0.679%

The CNN’s output percentages for the input airplane image for the most part fall in line with our expectations. The CNN seems to mistake an airplane for a ship, but that’s understandable, because they are relatively similar looking objects. It’s difficult to see how the intermediate images correlate to the final results, because the airplane and ship intermediate images don’t stick out significantly from the other intermediate class images.

1. Performance Evaluation

k-Plot results:

|  |  |
| --- | --- |
| k=1 | 0.4371 |
| k=2 | 0.6591 |
| k=3 | 0.7864 |
| k=4 | 0.8626 |
| k=5 | 0.9135 |
| k=6 | 0.949 |
| k=7 | 0.9710 |
| k=8 | 0.9847 |
| k=9 | 0.9942 |
| k=10 | 1 |

The algorithm was only correct 43.71% of the time. However, it gets to above 75% in the top 3 results, and above 90% in the top 5 results. It would be interesting to see what the numbers look like with more object classes, especially as the one in lecture had such high accuracy (AlexNet).

Confusion Matrix results:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | airplane | autom | bird | cat | deer | dog | frog | horse | ship | truck |
| airplane | 531 | 40 | 87 | 39 | 53 | 19 | 10 | 32 | 192 | 69 |
| automobile | 41 | 519 | 8 | 18 | 6 | 7 | 7 | 7 | 84 | 191 |
| bird | 65 | 9 | 386 | 127 | 270 | 151 | 120 | 73 | 35 | 23 |
| cat | 37 | 26 | 117 | 325 | 69 | 222 | 125 | 98 | 44 | 41 |
| deer | 10 | 10 | 97 | 45 | 259 | 49 | 93 | 77 | 7 | 4 |
| dog | 8 | 7 | 70 | 136 | 38 | 281 | 23 | 94 | 8 | 9 |
| frog | 18 | 19 | 104 | 186 | 162 | 111 | 557 | 54 | 10 | 30 |
| horse | 38 | 29 | 88 | 89 | 114 | 125 | 33 | 533 | 16 | 68 |
| ship | 210 | 111 | 25 | 13 | 22 | 20 | 9 | 13 | 542 | 127 |
| truck | 42 | 230 | 18 | 22 | 7 | 15 | 23 | 19 | 62 | 438 |

Green represents each time the highest probability matched the trueclass.

Blue represents the class each trueclass was “confused” for the most.

Generally, animals (bird, cat, deer, dog, frog, horse) were able to be distinguished from the transportation classes. The biggest difference in between highest value and second highest value was with frogs, which (subjectively) don’t look like the other animals in the list.

Airplanes and ships were confused with each other each way, as were trucks and automobiles, signifying that they are closely related in structure according to this CNN.

Dogs and cats were also closely related, being the most confused for each other.

1. Explorative Step

In this step we looked at an airplane and a boat that weren’t part of the test set. The results can be seen by navigating to /demos and running typing ‘demo1()’ and ‘demo2()’ into the console.

The airplane comes back as 54% ship and 42% airplane. As stated before, ships and airplanes were often confused with each other, so these results weren’t surprising.

The boat is 34% automobile, 25% ship, 17% airplane, and 23% truck. Again, ships and airplanes were confused with each other a lot as seen in the results, but it couldn’t figure out if it was an automobile/truck or ship/airplane. The quality of the picture wasn’t the best, but at least it knew it probably was not an animal.

For unknown objects, we could try putting in mixtures of plane/automobile (car with wings) and checking if the plane/automobile probabilities are the highest.

Additionally, we tested the CNN using a picture of one of the faces of our group members.

Input image (32x32x3):

demo/im.jpg

The output results from the softmax layer:

airplane 0.059%

automobile 0.687%

bird 0.594%

cat 16.594%

deer 1.015%

dog 1.423%

frog 77.551%

horse 0.931%

ship 0.431%

truck 0.717%

It seems the CNN reacts strongly to similar looking organic objects, just the same as it reacting strongly to mechanical objects. The CNN gave a high percentage match to a frog for this image, which we presume is due to the fact that the figure is wearing headphones – being mistaken as frog eyes.

1. Contributions

* Collective Development:
  + CNN framework
  + Normalization layer
  + ReLU layer
  + Quantitative Evaluation
* Lee’s Development:
  + Softmax layer
  + K-Plot
* Traister’s Development:
  + Fully Connected layer
  + Written report
* Scribano’s Development:
  + Convolution layer
  + Maxpool layer
  + Debug function
  + Confusion Matrix